Dr. Qu, lead discussant, was very impressed with the progress made by ICCVAM and NICEATM. She advocated strong collaborations and teamwork among several fields including statistics, computer science, biology, and toxicology. She said MRI data is much more accurate than just observing animals and she questioned how all the technology is being used for the 3Rs. She suggested emphasizing to Federal agencies the cost savings and ability to get accurate, quantitative data by using *in vitro* testing.

Dr. Niemi was impressed with ICCVAM's efforts but questioned using cells from 35 mouse strains. Dr. Tice explained that NTP's host susceptibility program is studying the relationship between genetic background and disease in the 35 strains. Using cells from those strains will assess whether the *in vitro* technology would identify differences in sensitivity based on a pathway analysis. The animal models would then be tested for high- and low-responders.

Dr. Corcoran said he found the ICCVAM Biennial Report very valuable and suggested adding two sections under each strategic goal, placing all work within the strategic context and closing with a strategic reflection. This should help with tying spending to strategic planning, prioritization, and increasing impact. He suggested convening biennially the key individuals in the 15 agencies. Drs. Stokes and Bucher agreed with the suggestion. Dr. Bucher said there was initially skepticism regarding HTS, so it wasn't mentioned in the FYP; however, it is moving forward and has a lot of promise for the prioritization of chemicals in the TSCA reauthorization. He said another area of HTS involves analysis to relate genes and disease pathways. Intersection of those two activities is critical because it all has to fit together to bring out the total potential of the HTS program. He agreed that relating the PK information to human blood levels is important in understanding toxicity pathways. Dr. Bucher expressed support for the NTP providing blood level information.

Dr. Brown agreed that the Biennial Report was valuable and asked about its distribution. Dr. Stokes said its publication is announced in the Federal Register, by trade organizations, and on various listserves. He welcomed suggestions for further distribution. Dr. Brown said many in the veterinary biological community are not aware of NICEATM and ICCVAM.

VIII. Current Issues in the Validation of Alternative Methods for Assessing Chemically-Induced Eye Injuries VIII. Current Issues in the Validation of Alternative Methods for Assessing Chemically-Induced Eye Injuries

A. Presentations

Dr. Stokes presented an overview of two technical issues that arose during a recent ICCVAM and NICEATM evaluation of alternative methods used to identify chemically-induced eye injuries. The issues are 1) The *Minimum Number and Proportion of Animals with Eye Injuries for Classification of a Chemical as an Eye Irritant*, and 2) *Reduced Eye Hazard Labeling Resulting from Using GHS Criteria Instead of U.S. Classification Criteria*. Dr. Stokes briefed the committee on the importance of eye safety testing and eye hazard labeling, and the larger context of the issues to be

discussed. Two million eye injuries occur annually in the U.S., representing a significant burden in terms of health care costs, lost workdays, and temporary and permanent disability. Chemicals and compounds are the third most common product category associated with eye injuries, accounting for 13% of all eye injuries, or an estimated 260,000 injuries annually. The EPA, CPSC, FDA, and OSHA require eye safety testing and labeling of potential eye hazards to provide safety messages to help prevent injuries.

Dr. Stokes reviewed the EPA's Eye Injury Hazard Categories and Labeling Requirements, which are based on the rabbit Draize test, which is currently the standard test method used for all worldwide eye hazard classification and labeling. Category I, labeled DANGER, involves severe eye damage and eye injuries lasting more than 21 days. Category II, labeled WARNING, involves injuries that clear within 8-21 days. Category III, labeled CAUTION, flags for injuries lasting seven or fewer days. There is also a Category IV, with an optional CAUTION label, involving injuries that resolve within 24 hours.

All regulatory hazard classification systems use the same scoring system for the nature and severity of lesions. However the classification criteria used to determine whether a chemical would require hazard labeling, and the appropriate hazard category, vary widely among U.S. agencies, nations, and international organizations. These classification criteria are based on the frequency, nature, severity, and duration of the eye injuries. The EPA has its own system, while CPSC and OSHA use a system based on the Federal Hazardous Substance Act (FHSA) regulations. A United Nations Globally Harmonized System for the Classification and Labeling of Chemicals (GHS) was originally published in 2003 and last revised in 2009. GHS is currently under consideration for implementation by U.S. agencies. ICCVAM evaluates new test methods for their accuracy for correctly classifying the hazard potential of chemicals for each of the U.S. and international hazard classification schemes. This involves calculating sensitivity, specificity, false positive rates, and false negative rates for each of the classification systems by comparing the *in vitro* predicted hazard category to the assigned hazard category in each system resulting from the *in vivo* reference test method.

However, two issues arose during a recent evaluation. First, ICCVAM encountered difficulty in assigning and classifying chemicals as eye hazards using FHSA classification criteria when it recently reviewed available *in vivo* reference data. This arose due to the fact that many chemicals would have required additional animal testing to assign a definitive FHSA hazard category, but such testing was not conducted. NICEATM, in consultation with the ICCVAM Ocular Toxicity WG, performed analyses to identify FHSA hazard classification criteria that could be used to classify these substances without additional testing, and criteria that could be used to classify substances when only 3 animals are used as recommended in several current test guidelines for *in vivo* ocular safety testing, instead of six to 18 as required in the current FHSA regulations.

The second issue NICEATM found during its analyses was that one *in vitro* method correctly identified chemicals that would not require eye hazard labeling using the GHS system, but failed to identify several chemicals as eye hazards that are currently classified and labeled as eye hazards in the U.S. Further investigation revealed a significant discrepancy between the GHS eye hazard criteria and current EPA, OSHA, and CPSC eye hazard classification criteria. The GHS criteria significantly reduce labeling of potential eye hazards compared to current U.S. criteria, with over 30% of chemical eye hazards no longer identified as hazards using GHS criteria.

Dr. Stokes reviewed the current *in vivo* rabbit test, including how rabbit eye injuries are scored in cornea, iris, and conjunctiva tissues. In the cornea, there is a 4-point scale for scoring positive lesions. In the iris, there are just two scores for positive lesions. In the conjunctiva, redness is scored as a 1 for minor redness, but only a score of 2 or 3 for increasingly severe lesions are considered as positive scores. Chemosis, or conjunctival swelling, is scored as a 1 for minor swelling, but only a score of 2-4 for increasingly severe lesions are considered as positive scores.

Dr. Joseph Haseman presented data regarding numbers of animals used in ocular testing. The FHSA regulations require a classification system involving up to three tests, each involving six animals. If the first test is inconclusive, there is a second test, and a third if the second is also inconclusive. Thus, up to 18 animals may be used with this approach. However, current best practices for eye irritation/corrosion tests normally use only up to three animals, so a comparison was needed to ensure that the smaller sample size would retain the appropriate sensitivity and specificity compared with the larger sample size tests, with the same level of hazard labeling as the current regulatory requirement (16 Code of Federal Regulations [CFR] 1500.42). His view was that the current FHSA sequential testing strategy is not very protective. He stated that he would show that a decision procedure based on just three animals is on average at least as protective as the current sequential procedure using up to 18 animals. Going over the sequential procedure, he showed that it contained some questionable aspects. For example, a positive response in 1 out of 6 test animals would be interpreted in three different ways in the three sequential tests, despite the fact that biologically the response is equivalent; in the first test, it's considered negative. In the second, it's considered inconclusive. In the third, one of six is considered positive, and the substance is labeled an eye hazard. He further showed that a positive interpretation, with labeling, could be generated by responses in as few as 4 of the 18 (22%) animals. In another scenario of the decision sequence, as many as 5 of 18 responses would result in no labeling. Dr. Haseman showed a chart depicting the number of animals required to assign an irritant classification under the sequential testing strategy. At the stage of the test, the minimum number of positive animals was four of 18, or 22%, but the maximum number of negative animals for a decision *not* to label was five of 18 or 28%. Ultimately, the sequential testing strategy appears to be confusing and may result in anomalous findings.

Dr. Haseman presented the results of his calculations comparing the sequential testing strategy with two versions of the three-animal strategy, one of which involves a positive

threshold of just one animal, the other requiring two or more positives for a decision to label. He called these Strategies 1 (sequential), 2, and 3. To effectively compare the protective value of the strategies, he looked at a range of underlying response rates. He found that Strategy 3 (which he identified as being roughly equivalent to the GHS system) would identify far fewer irritants than 1 or 2. Strategy 2, with its zero tolerance for positives, would be far more protective than the others, in that it would label more often based on lower underlying response rates. Ultimately, his study showed that Strategy 2, using a criterion of at least one positive animal in a three-animal test, would be at least as protective as current FHSA testing requirements, and that changing to that strategy would result in a saving of up to 83% fewer animals. Thus, he concluded, the three-animal strategy has much stronger basis for its use compared to the current sequential testing approach.

Dr. Stokes then presented the conclusions of the NICEATM analysis regarding Issue 2, the reduction in eye hazard labeling that would result from using the GHS criteria instead of U.S. criteria. NICEATM compiled and analyzed actual *in vivo* rabbit eye test results for a total of 262 chemicals from two databases, and calculated and compared EPA, FHSA, and GHS hazard classifications for each substance. Of 168 chemicals considered to be eye hazards under EPA classification criteria, 59 (35%) would not be so labeled as hazards under GHS. Of the 73 chemicals labeled as EPA Category III eye hazards, 57 (78%) would not be labeled under GHS, while two EPA Category II chemicals would not be labeled under GHS. Dr. Stokes presented data regarding the severity and duration of the eye injuries presented by the 59 chemicals that would not be labeled under GHS criteria. Forty-two % of the GHS "not labeled" chemicals produced grossly visible corneal and/or irideal injuries expected to interfere with normal vision. Twenty-five % of the chemicals had visible corneal and irideal injuries at 48 hours after application, and 19% had visible injuries at 72 hours post-application. Using the FHSA criteria, up to 30% of FHSA eye hazards would not be labeled as ocular hazards under GHS. Ultimately, using GHS criteria resulted in no hazard labeling for 30-35% of substances currently labeled as eye hazards under U.S. Federal regulations.

U.S. regulatory agencies are currently considering adoption of the GHS eye hazard criteria, and OSHA issued proposed rule making in 2009 to adopt the GHS criteria. Dr. Stokes emphasized that the GHS was negotiated with and emphasizes the principle that "the level of protection offered to workers, consumers, the general public, and the environment should not be reduced as a result of harmonizing the classification systems." However, there are no data to support that the reduced labeling for eye hazards that will result under GHS would not reduce the level of protection of workers and consumers provided by current U.S. regulatory hazard labeling.

He reiterated the main reasons why GHS criteria reduce eye hazard labeling compared to U.S. regulations. First, the minimum number and proportion of animals required to classify a substance as an eye hazard differs significantly, with GHS requiring that a minimum of two out of three animals must have positive responses, compared to only one out of three in U.S. requirements. Secondly, the GHS requires a greater severity of

eye injury as the minimum criteria for a positive response compared to the threshold for a positive response in U.S. requirements.

Dr. Stokes said there is a process for updating the GHS, which appears to be necessary to achieve hazard labeling that will support the GHS principle that the level of protection should not be reduced by the harmonization. Three proposals have been developed for optional or revised GHS labeling criteria that can provide hazard labeling at least equivalent to that provided by current U.S. regulations and therefore avoid the reduction in hazard labeling. These include: (1) adding an optional category for countries wishing to maintain their current level of hazard labeling; (2) retaining the current GHS criteria for Category 1 and 2A, but revising the current optional GHS Category 2B criteria to classify substances as ocular hazards based on positive ocular injury score in a least one animal (vs. the current two or more) at any of the three daily time points (vs. a three-day mean score), and (3) revising the current GHS Category 2 to classify substances as ocular hazards based on a positive ocular injury score obtained in at least one animal (vs. two or more) at any of the three daily time points (vs. a three-day mean). Any of the three proposals would identify all 59 EPA and FHSA eye injury hazards not currently classified by GHS.

Dr. Freeman asked Dr. Stokes how this particular issue is relevant to SACATM, because this agenda item does not appear to be concerned with validation or adoption of alternative test methods, but rather, is about the criteria for hazard classification of chemicals. Dr. Stokes said this was an issue more technical than the typical issues brought to SACATM for the group's input. However, as an *ad hoc* issue relevant to the evaluation of the validation status of *in vitro* test methods for regulatory safety testing, it was considered important to bring it to SACATM's attention, and to gain SACATM's perspective on the scientific analyses and questions involved. He emphasized that NICEATM and ICCVAM are not asking SACATM for a decision on whether GHS should be accepted, because the U.S. has already agreed to implement GHS, but rather to obtain SACATM's feedback on the appropriateness of the data analyses and conclusions. Dr. Freeman asked for clarification on the objective of the three proposals. Dr. Stokes said the proposals were drafted as three options for updating the GHS to allow for hazard labeling categories that could be used by the U.S. and other countries to maintain the same level of hazard labeling as currently required by their national safety regulations, consistent with the GHS principles. Dr. Corcoran asked about the EU system in relation to the systems used by EPA in the U.S. and by Health Canada, which are more protective. Dr. Stokes replied that there are no data available to assess the effectiveness and level of protection afforded by different national requirements, due to gaps in the eye injury reporting system currently in place. Dr. Corcoran clarified that he was interested in the animal testing data, which apparently is significantly less sensitive and protective under the EU system. Dr. Kreysa replied that GHS is actually somewhat more protective than the older EU system, which apparently had not resulted in any increase in eye injuries due to classification of substances as non-hazards, although he acknowledged that there is not a systematic monitoring system in place or other data to confirm this.

B. Public Comment

Karen Barlet, Monroe, North Carolina, shared her story of having been in a serious automobile accident in 2002. The deployment of the air bag in her vehicle resulted in her eyes being burned by the chemicals in the air bag propulsion system, leading to severe eye injuries. After many operations since the accident, she ultimately lost her left eye, and her right eye continues to deteriorate and would eventually also need to be removed. Although she had high praise for her caregivers, particularly Drs. Craig and Amy Fowler, she urged the committee to bring more attention to the danger of serious eye injuries associated with the chemicals in air bags, and the importance of warnings for consumers about the presence of chemicals that can cause eye injuries. She also advocated educating first responders to be aware of the danger, which would allow them to remove victims from vehicles more quickly and treat their potential eye injuries more effectively. Drs. Brown and Freeman expressed their sympathy to the speaker and thanked her for sharing her story.

C. SACATM Discussion

Dr. Proia, Professor of Pathology at Duke University and *ad hoc* discussant, praised Dr. Haseman's statistical analysis, and recommended strongly against adoption of the current GHS standards. Although the rabbit is not a good model for human injury, it is exquisitely sensitive, he said, and he would not embrace weakening standards.

Dr. Peiffer, veterinary ophthalmologist at Merck and *ad hoc* discussant, concurred with Dr. Proia that the Draize test is flawed, subjective and crude, but the only standard currently available. He opposes reducing the current requirement to a three-animal Draize test due to the potential for greater likelihood of false negatives. Dr. Proia cited the variability of eye injuries and among individuals, and the difficulty of defining clinical relevance. Dr. Peiffer said he found it especially concerning that the GHS system missed some EPA Category 2 compounds, and that there is no question that one would not want to expose one's ocular surface to one of those compounds.

Dr. Proia felt that it would be dangerous to remove hazard labeling from any of the currently labeled chemicals, as it would lead people to become more lax in their handling of the substances and would likely lead to more eye injuries as a result. Dr. Peiffer said there was a lack of human eye injury data, and urged that a mechanism be established to gather that information, possibly a reporting system with ophthalmologists, who would see many of the patients with such injuries. He said the trend should be toward a system that is more, rather than less protective, and he endorsed any of the three proposed updates to GHS on that basis.

Dr. Proia said that in his experience *in vitro* methods were extremely complex, and that we are likely decades away from developing effective assays to replicate *in vivo* situations accurately and reliably. Dr. Freeman responded that *in vitro* assays validated to assess eye irritancy exist today, but focus on the classification issues rather than seeking to address biological questions. Dr. Stokes added that existing assays are capable of predicting some of the substances that can cause irreversible eye injuries, and have value on that basis. Dr. Peiffer suggested prospective studies involving

improved animal models in comparison to *in vitro* approaches, although that would necessitate more animal use. Dr. Proia reiterated Dr. Peiffer's comment on the lack of human ocular injury data. Dr. Stokes explained that the question sought to determine whether there would be any value in seeking more detailed information using modern ophthalmic instruments, as opposed to the subjective observations currently in use. Dr. Peiffer said the Draize test could certainly be refined to make the data more valuable. Dr. Proia said histological correlates with the changes observed in the Draize test would be helpful. Dr. Peiffer suggested measuring corneal thickness, and Dr. Proia suggested confocal microscopy to detect cell death.

Dr. Corcoran, lead discussant, felt that population variance would make the three-animal test difficult to accept. He agreed with Drs. Proia and Peiffer that a zero tolerance policy was called for, but expressed some hesitation in terms of the costs involved and the burden of regulation. Being overly protective, however, would support the EPA's approach of assessing risks to the most vulnerable populations. He was concerned about confounding with existing databases, and said he would be more comfortable assessing the range of injury along with the ranges of other effects.

Dr. Hansen, lead discussant, concurred with much that had already been said, but pointed out that he believes the average consumer does not understand the difference between the words *danger*, *warning*, and *caution*, which are used in EPA eye hazard warning categories. He considered labeling addressing treatment as far more important than the classification categories.

Dr. Olson, lead discussant, said that he found the 33% positive rate in the three-animal assay to be acceptable, but that two blinded readings of the assay may be necessary to avoid bias. He agreed with Dr. Hansen's remarks about the importance of first aid information being included on labeling, with a note to seek professional help after exposure being added to the labeling of the more dangerous substances.

Dr. Qu, lead discussant, said there should be more concern about false negatives than false positives in this area. She suggested using both eyes in the rabbit test, which might remove the confounding presented by individual variation within a population.

Regarding the availability of human data, Mr. Wnorowski asked Dr. Fowle about an EPA database of adverse human effects from compounds in the market. Dr. Fowle replied that EPA had looked closely at that database as a potential source of human data in this area, but found it to be lacking in a variety of ways that rendered it unusable. Dr. Toth asked about the attention paid to reversibility in the Draize test, given that the eye is already damaged, regardless of whether the injury is reversible, and why these studies could not be terminated after injuries are observed. Dr. Stokes replied that earlier humane endpoints had been proposed to the peer review panel that met in 2009, but there had been some reluctance to adopt some of the earlier endpoints because some of the injuries might actually reverse within the 21 day observation period. However, he noted that in some cases, when permanent damage is unequivocal, this could be used to stop a test. Mr. Wnorowski, lead discussant, concurred, stating that very often in

laboratories, studies are terminated as soon as a very severe reaction is seen, regardless of when it occurs.

Dr. Meyer expressed concern that the discordance between GHS and current U.S. standards would continue as new compounds are introduced. Dr. Stokes said that GHS will be adopted, so that would not be an issue. Dr. Meyer asked if that meant a reduction of standards. Dr. Stokes said that this would occur if GHS were adopted as currently written. However, there is the opportunity for the U.S. to utilize GHS procedures to request updating of the GHS to add an optional category that could serve to negate the reduction in protection. Dr. Stokes explained some of the differences between the U.S. and GHS standards, particularly in terms of what would be considered positive responses in animals. He elaborated on ICCVAM's responsibility to provide scientific data and analyses to agencies that can assist them in determining whether new methods are effective in generating data that does not result in less protection than existing test methods. Dr. Fowle said the Harmonization Act refers to harmonization of test protocols, not classification schemes, which gets at policy and risk assessment. Dr. Fowle said it is not the role of ICCVAM to address classification schemes, only test protocols.

June 18, 2010
Dr. Freeman reconvened the meeting at 8:30 AM. Attendees introduced themselves and Dr. White read the conflict of interest statement.

IX. Updates on International Collaborations

A. European Centre for the Validation of Alternative Methods

Dr. Joachim Kreysa showed a promotional film made by the European Commission's Joint Research Center Institute for Health and Consumer Protection on protecting the European consumer and the use of science for a healthier life. The film demonstrated ECVAM's promotion of the development and dissemination of alternative methods to replace animal testing of consumer products. Dr. Sharon Munn, ECVAM coordinator, and other scientists describe the work done using human cells from umbilical cord blood, nanotechnology, advanced computational methods, high throughput systems, and the ECVAM databases.

Dr. Kreysa noted the assays for which validation is completed, but that are still in the regulatory acceptance process: the rLLNA, ICCVAM-ECVAM-JaCVAM harmonized LLNA Performance Standards (included in the revised OECD TG 429), three *in vitro* skin irritation tests (Epiderm™, Episkin™, and Skin Ethic™), the Guidance Document on using *in vitro* cytotoxicity tests to estimate starting doses for acute oral systemic toxicity, the Draft TG on *in vitro* micronucleus for genotoxicity, and the Guidance Document on application of the threshold approach for acute fish toxicity testing. TGs are in preparation by ECVAM for two cell-based assays for eye irritation, the Fluorescein Leakage Assay and the CM Assay.